REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave Blank*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 15, 1996	Final Report, 9/15/95–12/15/12

4. TITLE AND SUBTITLE
Thermal Analysis of Multichip Modules Using Domain Decomposition and Wavelet-Capacitance matrix

5. FUNDING NUMBERS
F49620-95-C-0073

6. AUTHOR(S)
M. An, R. Tolimieri, J. Weiss, J. Byrnes

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Prometheus, Inc.
103 Mansfield Street
Sharon, MA 02067

CCNY/CUNY and RF CUNY
Convent Avenue and 138th Street
New York, NY 10031

AFOSR-TR-97
0299

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFOSR
110 Duncan Avenue, Suite B115
Bolling AFB, DC 20332-0001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release;
distribution is unlimited.

12b. DISTRIBUTION CODE

A

13. ABSTRACT (*Maximum 200 words*)

Computational technology base for the solution of the initial value problem for 2-dimensional heat diffusion equation in a finite, composite multidimensional region has been developed. This is the first step toward creating a software design tool whose purpose is to acuurately predict reliability and durability of multichip module designs. Singularities arising from material interfaces have been resolved by the multiresolution analysis based on wavelet-Galerkin method. Implementation strategies used for the wavelet-Galerkin method are: domain decomposition for partitioned and distributed computations; incorporating local and global data symmetries in all stages of the computations for reduction of memory and computations; using algebraic formalism for algorithm development and code generation; incorporating optimized FFT routines. Codes have been tested on several hardware platforms, both single and multiple CPU's.

14. SUBJECT TERMS			15. NUMBER OF PAGES
thermal analysis	domain decomposition	FFTs	10
wavelet-Galerkin	software design tool	symmetries	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	None

1 Introduction

Multichip modules (MCMs) incorporate several microcircuits into a single package, offering the potential to increase circuit density to over 100 chips per module package. This technology was first proposed in the Wafer Scale Integration (WSI) Program at Rome Laboratory. There are now commercial vendors, including Mercury Supercomputer Systems, Inc., Intel Corporation and Honeywell Incorporated, developing and supplying MCMs. There are varied implementations of WSI, but with the same goal of increasing the electronic circuit density, the reliability and the module performance. While MCMs hold potential for delivering the computational density required for timely processing of large amounts of data, the reliability assessment of MCMs is a challenging task due to the high complexity [1] of modeling the behavior in inhomogeneous multi-material media. Material interfaces are a major source of singularities. Inherent complexity, high unit cost and low production levels make traditional statistical based methods of testing impractical. The solution is to develop the reliability assessment process during the design stages. An important component of this effort is the finite element thermal and stress analysis of the multichip design. In fact, this is the critical component in terms of speed and accuracy. Since flaws in a design tend to occur at material interfaces, the accurate resolution of the thermal diffusion and induced stress effects at these singular interfaces is of great importance.

We have carried out a combined research and implementation project to evaluate the feasibility of our solution and implementation methodologies for creating a software tool for analyzing possible failures in the design stages of MCMs. One result of this effort is the development of a computational technology base for the solution of the initial value problem for the 2-dimensional heat diffusion equation in a finite, composite multimaterial region. Within this computational framework, we have also developed and implemented 3-dimensional computational cores based on FFT routines incorporating global data symmetries. Implementations were made on single processor Pentium, 4 parallel processor i860 and the multichip P6.

We have made the following technology transfers.

- Dr. John Hines of Wright Laboratory has agreed to beta test our existing and future codes for the purpose of insuring relevance and usability.
- We gave an invited presentation on our work at the University of North Carolina conference on "Wavelets, relations with operators and applications," held 24-28 July.
- Visited engineers (D. Holzhauer et al) of Rome Laboratory at Griffiss AFB. We discussed the capabilities and limitations of the existing software, IMCMA (Intelligent Multichip Module Analyzer), developed at Griffiss AFB. We have thus established a working relationship with Mr. D. Holzhauer et al. The significance of this relationship was two fold: IMCMA provides validation of our methods and codes; IMCMA is unique in its capability to address the currently most important geometry of flip chip configuration.
- Presented our results to scientists (Mr. Lyke et al) at Phillips Laboratory, Kirtland AFB.

The persons supported by and contributed to this project are Drs. M An, J. Weiss and J. Byrnes of Prometheus and Professor R. Tolimieri and Mr. D. Wahl of CCNY/CUNY.

2 Solving the Heat Diffusion Equation.

We have used an adaptation of finite element analysis (FEA) procedure incorporating wavelet bases. Being both compactly supported and orthogonal, wavelets combine the advantages of finite element, splines and Fourier spectral methods. Due to active research in the theory of wavelets, there now exist methods for exact evaluation of functionals of compactly supported wavelets required for correct application of the Galerkin procedures.

The time-dependent, heat diffusion equation,

$$(D_t - \bigtriangledown \cdot k \bigtriangledown + d)\psi = P,$$

describes the evolution of the temperature ψ in some region of interest. For multicomponent systems, k and d are, in general, discontinuous at the interfaces of components. The (internal) boundary conditions at an interface are [3]

$$k_- \psi_{\hat{n}}^- = k_+ \psi_{\hat{n}}^+ = h(\psi^- - \psi^+).$$

At an interface there is a jump in temperature ψ and its normal derivative, $\psi_{\hat{n}}$.

2.1 Compactly supported wavelets

To describe the class of compactly supported wavelets briefly, let ϕ be a solution of the *scaling* relation

$$\phi(x) = \sum_{k=0}^{N-1} a_k \phi(2x - k).$$

The a_k are a collection of coefficients that categorize the specific wavelet basis. The expression ϕ is called the scaling function.

The translates of ϕ are required to be orthonormal

$$\int \phi(x - k)\phi(x - m) = \delta_{k,m}.$$

From the scaling relation this implies the condition

$$\sum_{k=0}^{N} a_k a_{k-2m} = \delta_{0m}.$$

For coefficients satisfying the above two conditions, the functions consisting of translates and dilations of the scaling function, $\phi(2^j x - k)$, form a basis for square integrable functions on the real line.

If only a finite number of the a_k are nonzero then ϕ has compact support.

The compactly supported wavelet is defined by the equation

$$\psi(x) = \sum (-1)^k a_{1-k} \phi(2x - k).$$

The translates of the scaling function and wavelet define orthogonal subspaces, i.e.

$$\int \phi(x)\psi(x - m)dx = \sum (-1)^k a_{1-k} a_{k-2m} = 0.$$

2.2 The wavelet-Galerkin method

For a PDE of the form

$$F(U, U_t, \cdots, U_x, U_{xx}, \cdots) = 0$$

define the wavelet expansion

$$U = \sum U_k \phi(x - k).$$

An approximation to the solution is defined by

$$\hat{U} = \sum_{k=-M}^{N} \hat{U}_k \phi(x - k).$$

Thus the solution is projected onto the subspace spanned by

$$\Phi(M, N) = \{\phi(x - k) : k = -M, \cdots, N\}.$$

Herein and in what follows, we assume, for simplicity and without loss of generality, that the dilation factor 2^j has been normalized to 1 by a scale transformation $y = 2^j x$. In effect, the integers are the finest scale. To determine the coefficients of this expansion we substitute into the equation and again project the resulting expression onto the subspace $\Phi(M, N)$. This uniquely determines the coefficients U_k.

The projection requires $\hat{U}_k$ to satisfy the equations

$$\int_{-\infty}^{\infty} \phi(x - k) F(\hat{U}, \hat{U}_t, \hat{U}_x, \cdots) dx = 0$$

for $k = -M, \cdots, N$. To evaluate this expression we must know the *connection coefficients* of the form

$$\int \phi(x) \phi_x(x - k_1) \cdots \phi_{xx}(x - k_2) \cdots dx.$$

A typical functional (three term connection coefficient) is

$$\Omega(k, j) = \int \phi_{xx}(x) \phi_x(x - k) \phi(x - j) dx.$$

Since the scaling function used to define compact wavelets has a limited number of derivatives, the numerical evaluation of these expressions is often unstable or inaccurate. We have found methods for evaluating the functionals *exactly* based on the *scaling relation*

$$\phi(x) = \sum_{k=0}^{N} a_k \phi(2x - k).$$

By straightforward manipulations, a system of equations is found for the $\Omega(k, j)$. The system of equations is generally rank deficient (singular). The rank deficiency is cured and a unique solution is obtained by the inclusion of an additional set of linear equations that are obtained from the *moment* equations. The resulting system is non-singular and non-homogeneous and has a unique solution that is easily found by standard techniques. One of these techniques is derived in [4].

2.3 Domain decomposition

The wavelet-Galerkin and capacitance matrix methods for domain decomposition have been carried out by an implicit time differencing and wavelet-Galerkin discretization with the capacitance matrix method to impose boundary conditions.

For the wavelet-Galerkin method we expand k, d, and ψ^m in scaling function expansions

$$k = \sum\sum k_{i,j}\phi(x-i)\phi(y-j),$$

$$d = \sum\sum d_{i,j}\phi(x-i)\phi(y-j),$$

$$\psi = \sum\sum \psi_{j,k}\phi(x-j)\phi(y-k)$$

and apply the Galerkin procedure to determine the coefficients $(\psi_{j,k})$. To resolve the discontinuities of k and d we use the wavelet transform to find the smooth (large scale) and discontinuous (small scale) parts

$$k = k_s + k_i$$

$$d = d_s + d_i.$$

In effect, we expand (k, d) into scaling function and wavelet components. The discontinuous (wavelet) (k_i, d_i) are localized in a neighborhood of the interface. This process decomposes the diffusion operator into smooth and discontinuous parts

$$L = \bigtriangledown \cdot k \bigtriangledown + d = L_s + L_i.$$

With this representation the implicit time differencing is defined as

$$(I - dtL_s)\psi_{n+1} = (I + dtL_s)\psi_n + dtL_i(\psi_{n+1} + \psi_n)/2 + P$$

and solved by iteration at each time step.

Implicit time differencing is used to time advance the Galerkin coefficients. The internal boundaries cause singularities that are represented by a capacitance matrix term defined at the boundary. This is similar to the approach of [3] where a fictitious layer is used to represent imperfect contact at interface. Therefore, the interfaces are approximated as singularities using the wavelet-Galerkin approximation of the Green's function, desingularizing the approximation.

3 Digital Implementations

Formulating the implicit time differencing as a convolutional product, solution is found as the Fourier transform of the diffusion operator. By the convolution theorem, the convolution product is computed as the pointwise product of the Fourier transforms. This requires the computation of the 2-dimensional Fourier transform and its inverse for every iteration at each time step. Thus, over 90% of the computational burden is carried out by the 2-dimensional FFT.

3.1 Optimized FFT routines

There exists optimized FFT codes on wide ranges of hardware architectures. In fact, we have been developing and implementing many of these codes on distributed and parallel architectures based on special and general purpose digital signal processing chips and reduced instruction set computers. The availability of optimized FFT codes which act on highly flexible input and output data structures was of critical importance for efficient interfacing of the Fourier transform with current modeling strategies.

The 2-dimensional $N \times N$ DFT is defined by

$$y(n_1, n_2) = \sum_{k_1=0}^{N-1} \sum_{k_2=0}^{N-1} x(k_1, k_1) e^{-2\pi i (n_1 k_1 + n_2 k_2)/N}, \quad 0 \le n_1, n_2, k_1, k_2 < N.$$

Both the diffusion operator and data at each time step are real valued. Thus the memory (or cache) required is exactly $1/2$ that of complex data. Algorithm is given to show that the required computation is also roughly $1/2$.

Algorithm for real FT

- Define a complex 2-dimensional array u of size $N \times (N/2+1)$ from $x(k_1, k_2)$ by

$$u(k_1, k_2) = x(k_1, 2k_2) + ix(k_1, 2k_2 + 1), \quad 0 \le k_1, k_2 < N/2.$$

- Compute the N-point FT along the first dimension of u.
- Matrix transpose the result using the intermediate symmetry to the array v of size $N \times (N/2+1)$ from u.
- Compute the N-point FT along the second dimension.
- Define a complex 2-dimensional array y by

$$\begin{aligned} y(n_1, 2*n_2) &= 1/2\,(v(n_1, n_2) + v^*(N - n_1, N - n_2)), \\ y(n_1, 2*n_2 + 1) &= -i/2\,(v(n_1, n_2) - v^*(N - n_1, N - n_2)), \\ & \quad 0 \le n_1 < N, \;\; 0 \le n_2 < N/2. \end{aligned}$$

This algorithm as well the one below were implemented with minimum of 2-fold speed up over complex-to-complex FFT routines: smaller size data more than made up for the overhead in incorporating the symmetry. For computation sizes that are well over the cache of the machines, this routine performs many times (up to 20) faster than complex-to-complex routines by avoiding memory swapping.

For the inverse FT, we observed that the Fourier transform of a real sequence is Hermitian symmetric, i.e.,

$$y(N - n_1, N - n_2) = y^*(n_1, n_2), \quad 0 \le n_1, n_2 < N.$$

Thus again, the required memory and, as we will see, computation required are roughly $1/2$.

Algorithm for Hermitian FT

- Compute the N-point inverse FT along the first dimension of y.
- Define a complex 2-dimensional array u of size $N \times (N/2+1)$ from the transpose of $y(k_1, n_2)$ by

$$\begin{aligned} u(k_1, n_2) &= y(2 * k_1, n_2) + iy(2 * k_1 + 1, n_2), \\ u(k_1, N - n_2) &= y^*(2 * k_1, n_2) + iy^*(2 * k_1 + 1, n_2), \\ & \quad 0 \leq k_1 < N/2, \;\; 0 \leq n_2 < N/2. \end{aligned}$$

- Compute the N-point FT along the second dimension of u.
- Extract the desired result by

$$\begin{aligned} x(2 * k_1, k_2) &= real(u(k_1, k_2)) \\ x(2 * k + 1, k_2) &= imag(u(k_1, k_2)). \end{aligned}$$

Similar algorithms were developed and implemented for 90°-rotational symmetry with comparable performances.

4 Results and Conclusions

The goal of the proposed software tool is **computational efficiency without trading off numerical accuracy**. The critical issue determining the applicability of the software tool is portability and scalability of its implementation so as to interface with already existing technology. To this end, we have investigated and concluded that

- the wavelet-based numerical methods are effective.
- incorporation of symmetry is effective in terms of accuracy, computational complexity and memory requirements.
- computational and communication complexity of our numerical methods for optimal implementation is feasible.
- many hardware platforms are feasible for carrying out the required computations.

Our conclusion is based on the following.

- We have

  Designed, implemented and tested the wavelet-Galerkin solver for the time-dependent, two-dimensional and composite, heat equation

  Developed and applied an adaptive implicit time differencing for the time-dependent, two-dimensional and composite, heat equation that imposes a monotone convergence.

  Developed a new variant of a higher order, implicit Runge-Kutta time differencing that converges faster and with a time step that is two to three larger than the implicit Euler time differencing used previously.

Coded and optimized an alternate formulation of the wavelet-Galerkin solver for the time-dependent, two-dimensional and composite, heat equation that allows fast (FFT based) algorithms through a consistent wavelet-Galerkin operator splitting.

We have found that

- Continuous adjustment of the time step can control the (iterative) implicit time step error to be at a prescribed level.
- For a piecewise constant coefficient of diffusion there exists a fast (FFT based) implementation of the full wavelet-Galerkin method. The timings are: the full wavelet-Galerkin evaluation procedure is 95% of the program CPU run time. By fully taking into account the local tensor product and piecewise constant structure of the coefficient of diffusion, the cost of the exact wavelet-Galerkin evaluation can be reduced to about 5% of the program CPU run time (for the non-FFT evaluations).
- Variations of the time stepping method can increase the time step size and improve convergence for implicit time differencing schemes.

- We have applied optimized, multidimensional, parallel FFT algorithms to implement the convolutions and deconvolutions required by the wavelet-Galerkin methods.

  Implemented 2-dimensional real convolution kernels based on higher radix FFT incorporating global symmetries. Performance of the kernel is minimally 2 fold but often much more than that of radix 2 FFT.

  Designed 3-dimensional real convolutional kernels that can incorporate symmetries for reducing computation and memory requirements. These convolutional kernels are also denser than powers of two.

- We have identified computational cores that have been optimally implemented as well as those which require optimal implementation.

  Denser than powers of two FFT routines must be implemented. The density will be most important to keeping the problem sizes to be minimal required in higher dimensions.

  We have formulated a vector-radix multidimensional FFT algorithm [2, 5] for better adapting the mesh data decomposition structure. Mesh data decomposition is a more natural and efficient domain decomposition method for parallelizing our code. Compatibility of the vector-radix algorithm with various possible symmetries was one of the parameters in our formulation.

- We have begun the development of multichip module benchmarks for differing input geometries.

  This benchmark was tested for a two-component system separated by a narrow gap on a low diffusion background. The calculation is nearly as fast as that for the single component system done previously.

References

[1] D.J. Holzhauer, D.W. Richards, P.J, Rocci, M.J. Stoklosa, P.S. Yaworsky, I.R. Grosse, and D. D. Corkill, "Designing reliability into multichip modules: An intelligent modeling approach," Rome Laboratory Technical Report, presented at the second annual Aerospace Design conference, Feb. 1993. *SPIE real time signal processing proceedings* San Diego, 132-146, 1989.

[2] An, M., Gertner, I, Rofheart, M. and Tolimieri, R. "Discrete fast Fourier transform algorithms; A tutorial survey," *Advances in electronic and electron physics*, P. Hawkes, Ed. Academic Press, 1990.

[3] M.D. Mikhailov and M.N. Ozisik, *Unified analysis and solutions of heat and mass diffusion*, Dover Publications, Inc., New York (1994)

[4] A. Latto, H. L. Resnikoff and E. Tenenbaum, "The evaluation of connection coefficients of compactly supported wavelets," Aware Technical Report AD910708 (July 1991) and the Proceedings of the French - USA Workshop on *Wavelets and Turbulence*, Princeton University, June 1991, Editor, Y. Maday, Springer - Verlag (1992).

[5] Tolimieri, R., An, M. and Lu, C. *Mathematics of multidimensional Fourier transform algorithms*, Springer-Verlag, New York, 1993.